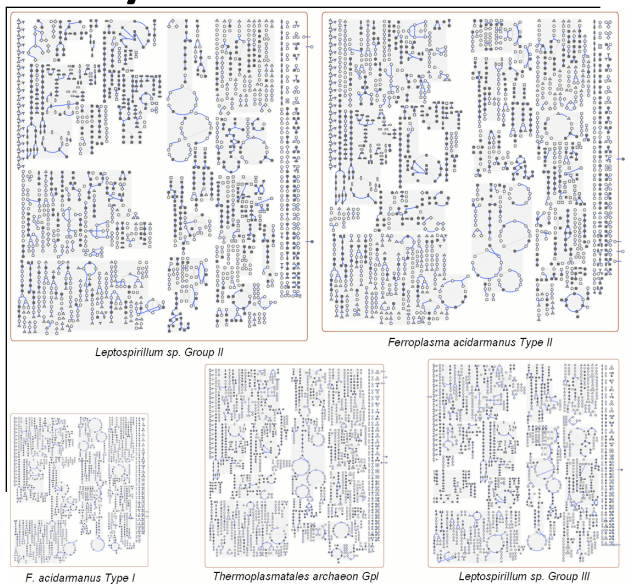
Genomics/MetaGenomics Populations/Communities/Ecosystems

Patrik D'haeseleer
Comp. Scientist / Systems Biology
LLNL, JBEI



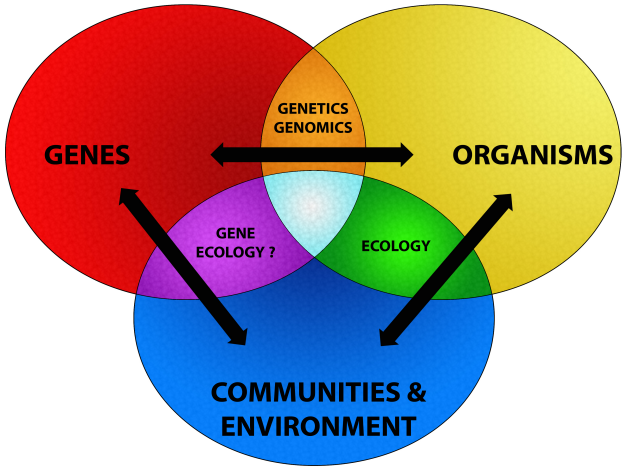
Reconstruction of the Acid Mine Drainage community. Metabolic networks for the five dominant species were drawn proportional to their metagenome coverage

Current Area of Research Interest

- Targeted discovery of enzymes and pathways
- Inference of metabolic networks and phenotypes from 16S, and noisy/incomplete metagenome data

Challenges that May be Addressed with Advanced Comp. & Math Capabilities

- Binning by probabilistic reconstruction of phylogeny of each metagenomic read
  - Already feasible for marker genes, or from oligonucleotide patterns in long sequences
- Bayesian reconstruction of metabolic network for each member of the community
  - Based on uncertain binning, <1X coverage, and uncertain enzyme annotations
- Agent-based model of interacting metabolic networks
  - Simulate 1000's of microbial cells, each with its own metabolic flux model

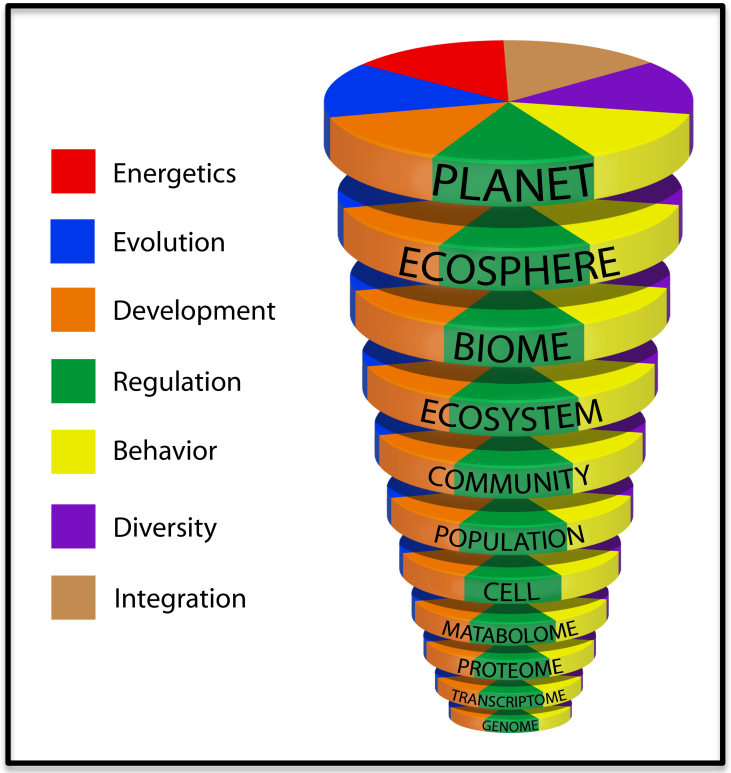


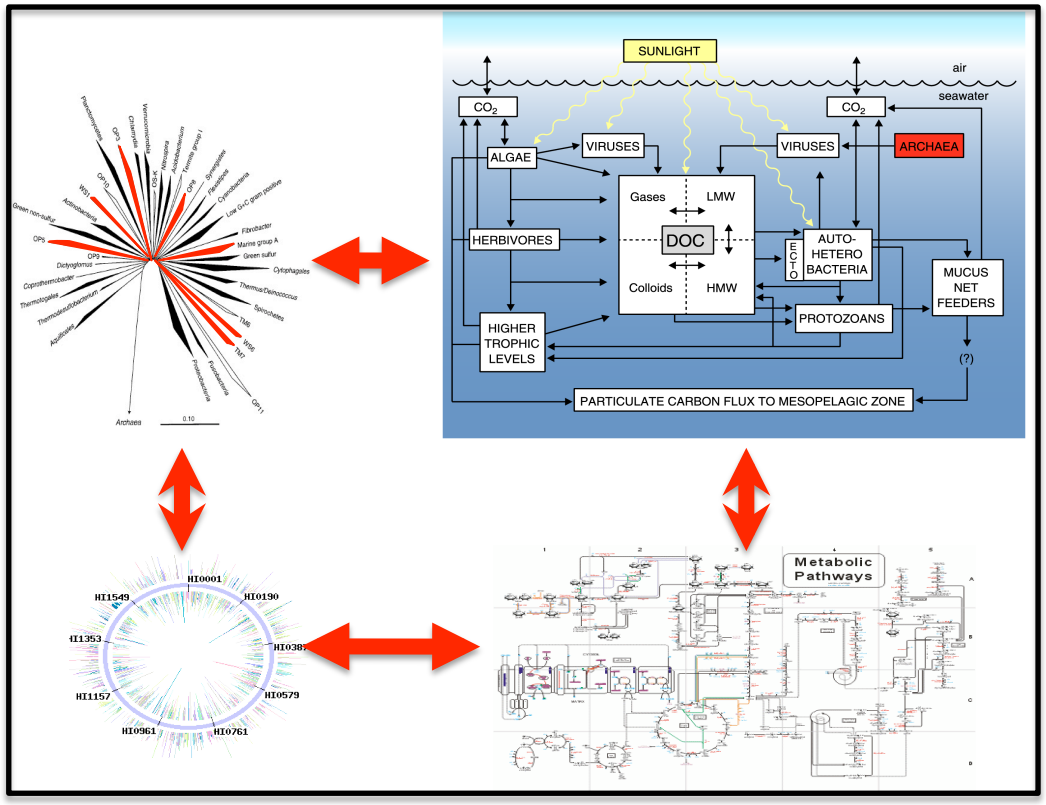
Systems biology on a planetary scale – some grand challenges
Ed DeLong

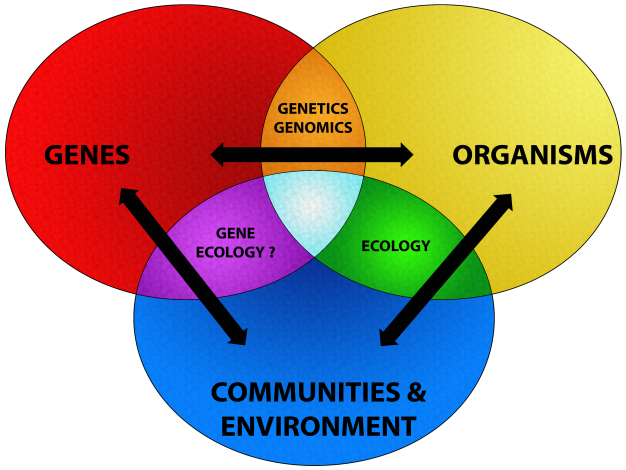
• Integrating measurement and modeling across multiple levels of biological structural organization

• Integrating measurement and modeling of biological process across appropriate spatial and temporal scales

• Integrating above biological structure/process at various scales with dynamic physical, chemical, & geological models





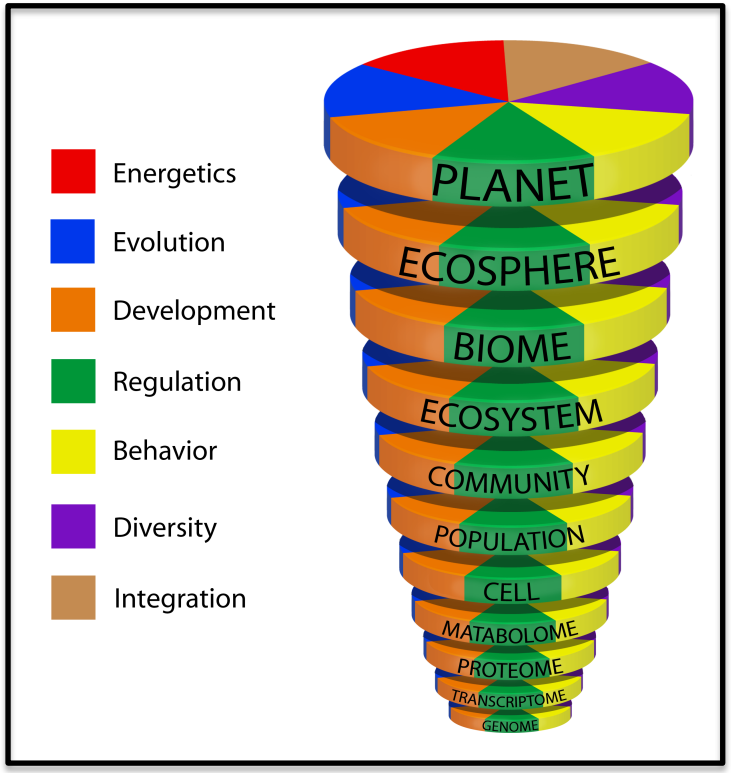


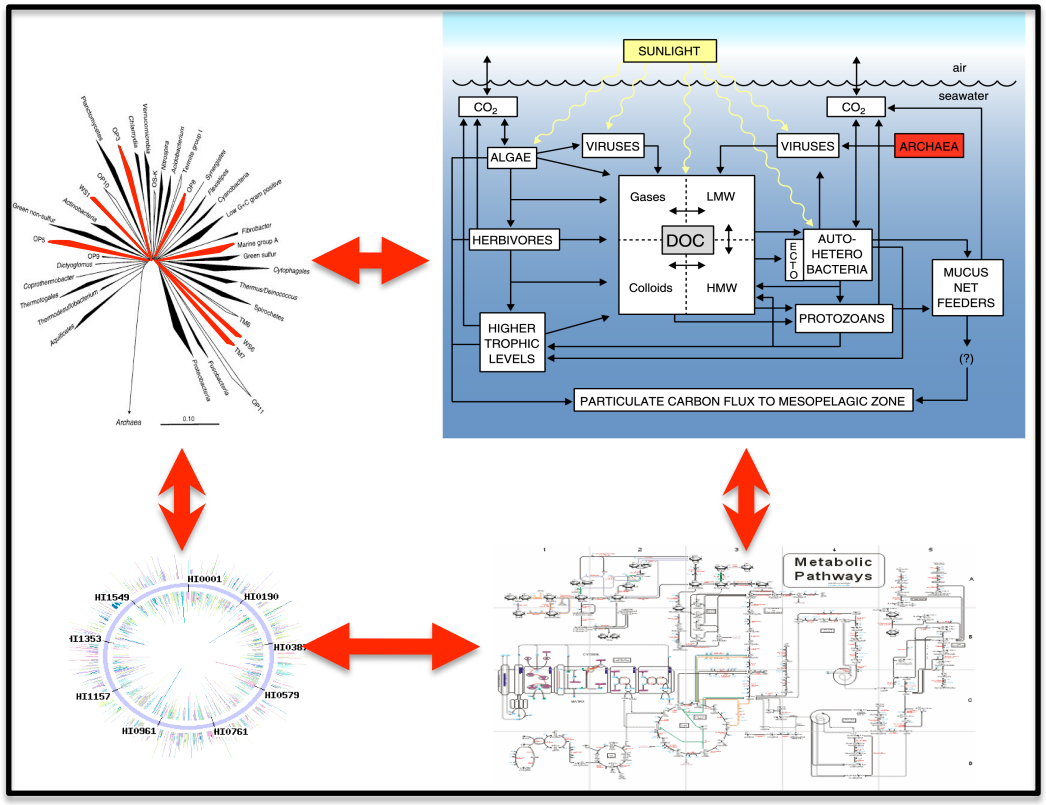
Systems biology on a planetary scale – some grand challenges
Ed DeLong

• Integrating measurement and modeling across multiple levels of biological structural organization

• Integrating measurement and modeling of biological process across appropriate spatial and temporal scales

• Integrating above biological structure/process at various scales with dynamic physical, chemical, & geological models



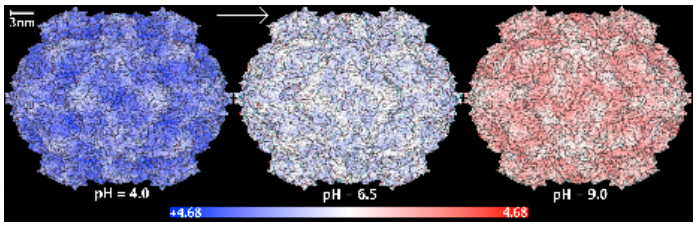


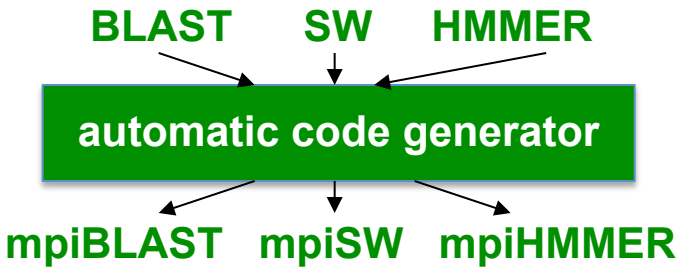
Populations, Communities, Ecosystems, and Evolutionary Dynamics: Genomics and Metagenomics

***Wu Feng**, Associate Professor*
Virginia Tech

http://www.mpiblast.org/
http://synergy.cs.vt.edu/

5000x speed-up on a GPU





Current Areas of Research Interest

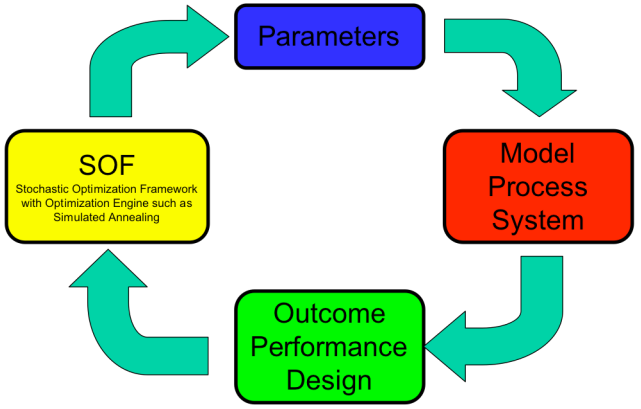
- High-Performance Biocomputing
  - ✓ Sequence Alignment
  - ✓ Framework for Automatic Parallelization of Biological Codes
  - ✓ Biomolecular Electrostatic Potential → Long-Range Interactions (w/ A. Onufriev)

Challenges Addressed with Advanced Computing & Mathematics Capabilities

- Sequence databases now growing at a rate faster than we can compute on them.
  - ✓ Accelerate algorithms in hardware & software, automate code generation, simplify maintenance
- Finding "Missing Genes in Genomes"
  - ✓ More complete picture of organism capability → help control disease, understand its metabolism to improve efficiency, improve knowledge of protein-coding genes in nature (e.g., new enzymes to more efficiently convert sugar to fuel w/ P. Zhang)

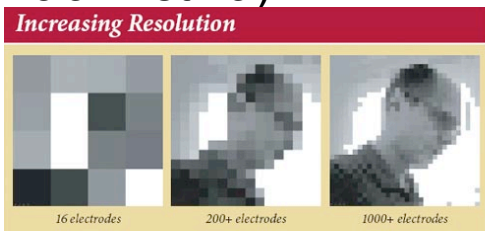
Populations, Communities, Ecosystems and Evolutionary Dynamics

Wolfgang Fink, Ph.D.
Senior Researcher
Caltech/JPL/NASA



Current Area of Research Interest [Include graphic]

- Optimization of Neuroprosthetic Devices (e.g., Artificial Retina)
- Autonomous Exploration Systems
- Image Processing and Analysis
- Stochastic Optimization Frameworks



Challenges that May be Addressed with Advanced Computing and Mathematics Capabilities

- De Novo Protein Design
- Drug Design
- Chemical Reaction Networks
- Cellular Pathways

Genomics and Metagenomics

George M. Garrity
Professor, Michigan State University

SIGS Standards in Genomic Sciences
An Open Access Journal of the Genomic Standards Consortium

NamesforLife™
Bringing meaning to life ...

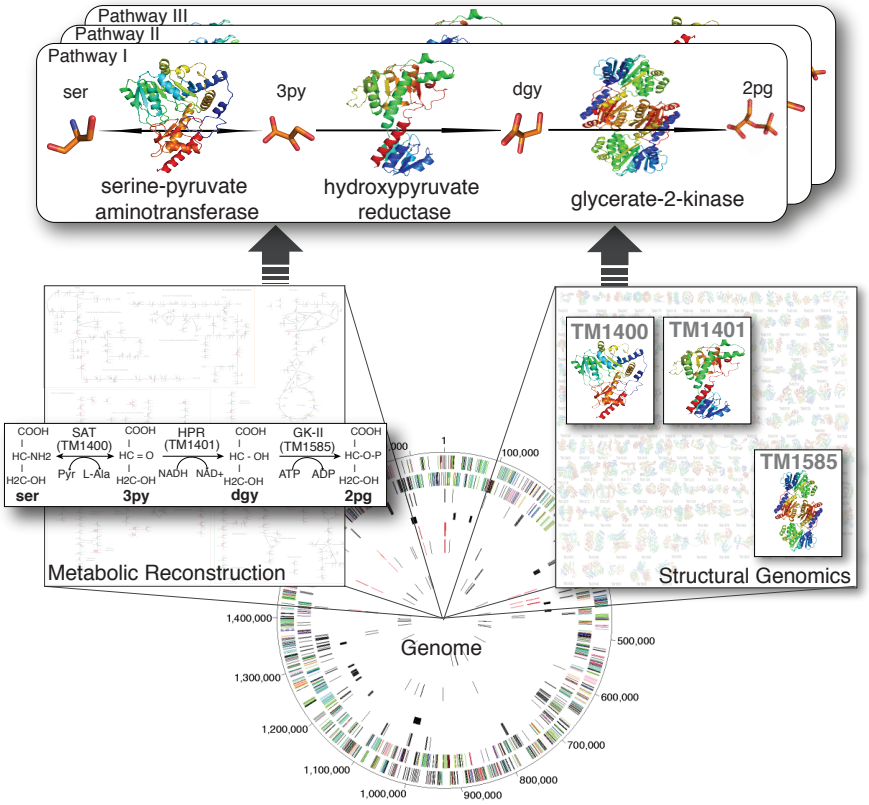
Current Area of Research Interest

- Systematics of *Bacteria* and *Archaea*
- Semantics and biological nomenclature
- Mining scientific, technical, regulatory and patent literature
- Metadata standards development and application

Challenges that May be Addressed with Advanced Computing and Mathematics Capabilities

- Extreme-scale phylogenetic/phylogenomic modeling of *Bacteria* and *Archaea*
- Extreme-scale clustering and visualization of natural patterns in biological data
- Improved standards for interoperability and exchange of data

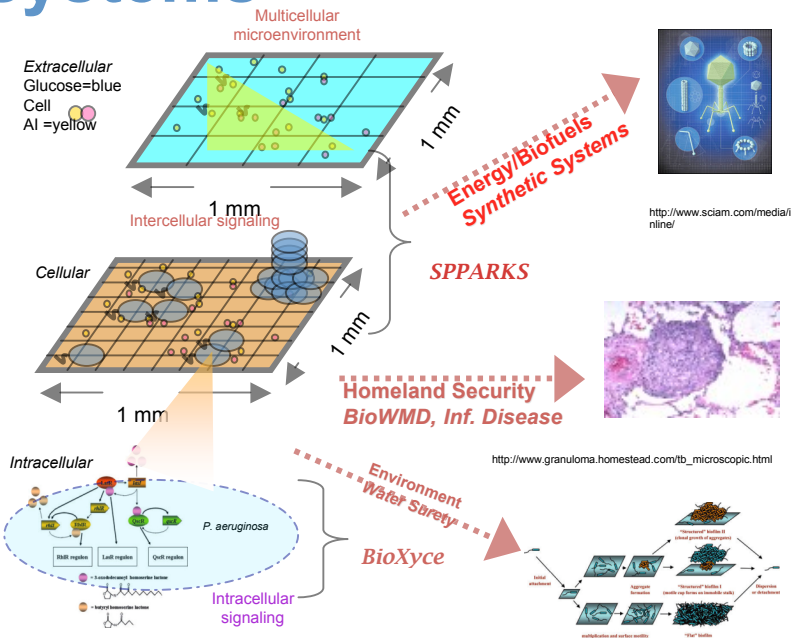
Structural systems biology: integrated structural and functional view of protein networks



- A group of proteins with limited number of topologies is responsible for a vast majority of functions (core machinery of life) and we are almost done cataloging their structures
- Using tools of distant homology recognition we can add at least low resolution structural information to functional networks of proteins (here, central metabolic network of thermophilic bacteria, *T. maritima*)
- Exploring synergy between function- and structure- centric view of the network we can gain new insights into the evolution, internal structure and other aspects of such networks

Populations, Communities, Ecosystems and Evolutionary Dynamics: Genomics and Metagenomics

Elebeoba May, PhD
Principal Member Technical Staff
Sandia National Laboratories



Current Area of Research Interest

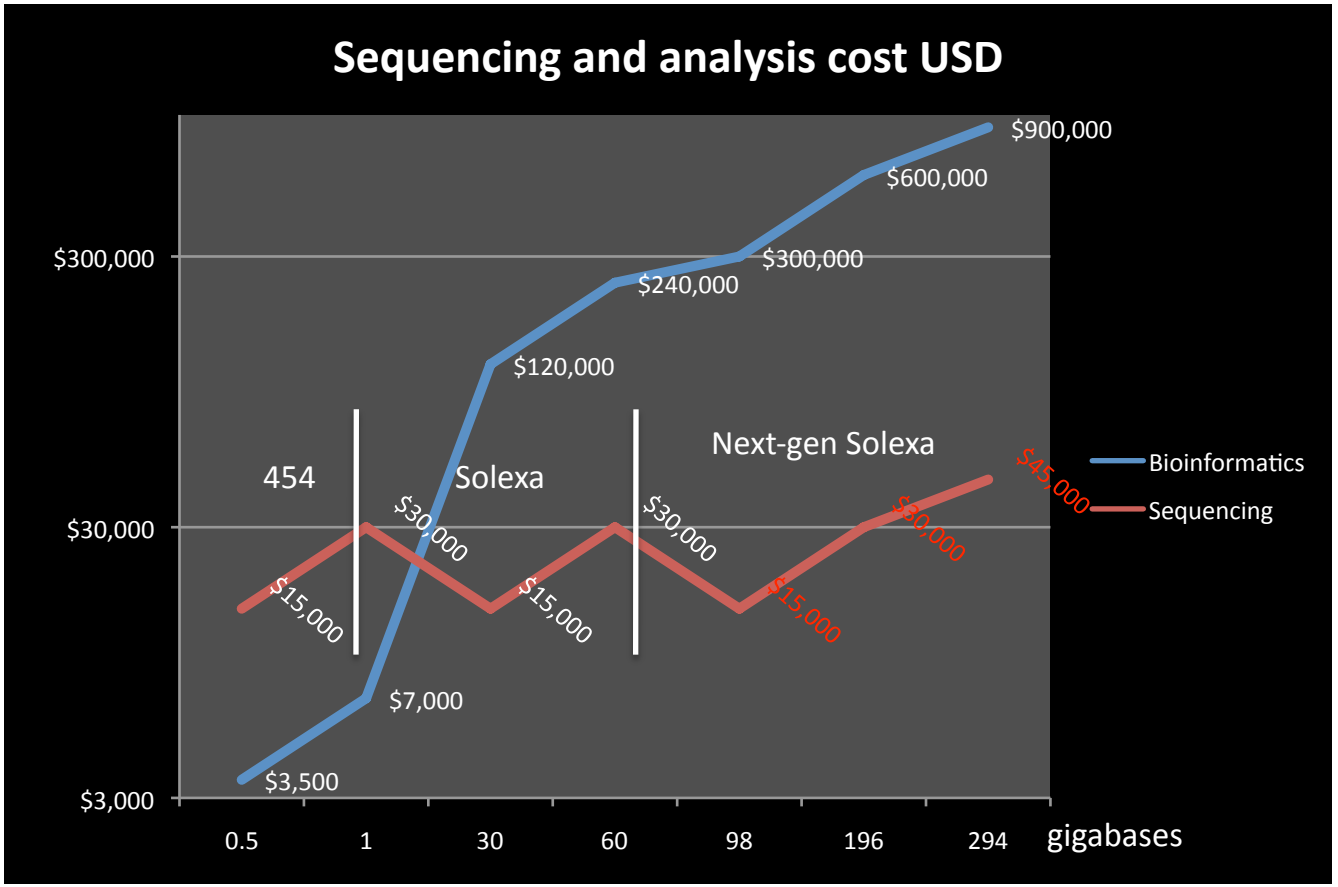
- Multiscale models of microbial communities and phenomena
- Modeling and simulation of host-pathogen interactions in infection and immunity
- Modeling and control of molecular communication

Challenges that May be Addressed with Advanced Computing and Mathematics Capabilities

- Accurate implementation of multiscale, spatio-temporal models of molecular systems and microbial communities (multi-platform simulation, large-scale data communication)
- Large-scale parameter estimation and optimization
- Large-scale reachability for analysis and ultimately engineering of microbial systems
- Multiscale visualization of biological systems and biological information

Folker Meyers Slide

The problem: Sequencing outpaces Moore's law

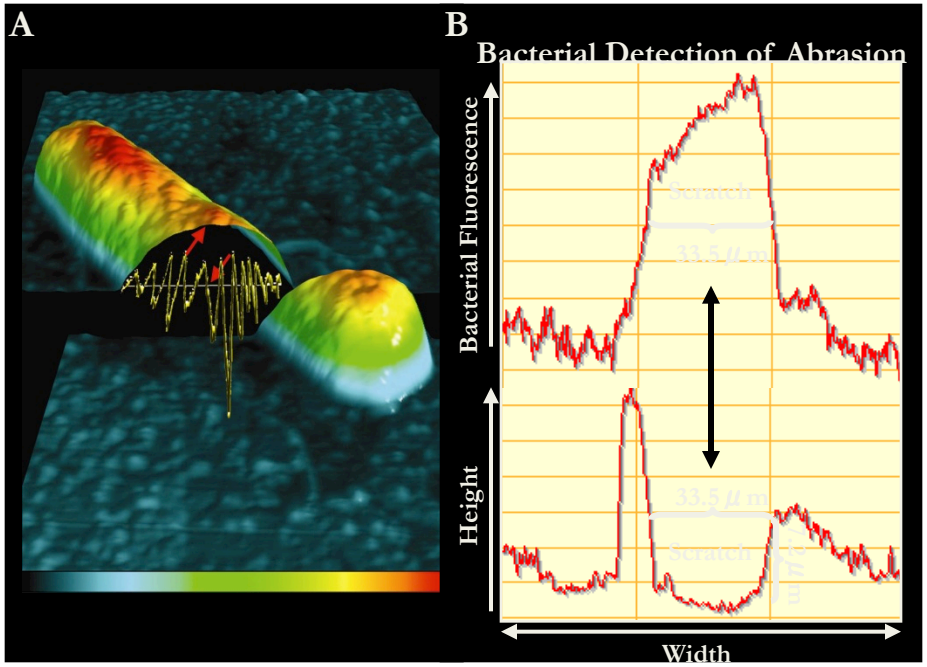


- Computational analysis of metagenomic data is and will be a very big challenge
- Bioinformatics cost is purely BLAST, reality is even worse
- Cost are in Amazon EC2 (from *Wilkening et al, IEEE Cluster09*)

Genomics and Metagenomics:

Populations, Communities, Ecosystems and Evolutionary Dynamics

Ken Nealson, Professor
U. S. Cal. - Earth Science
J.C. Venter Inst. - Biology



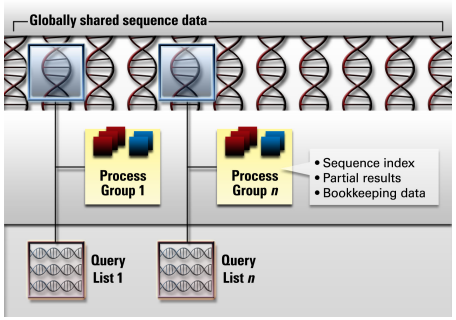
Current Area of Research Interest

- Environmental Microbiology
- Extracellular electron transport
- Energy flow in communities
- Corrosion

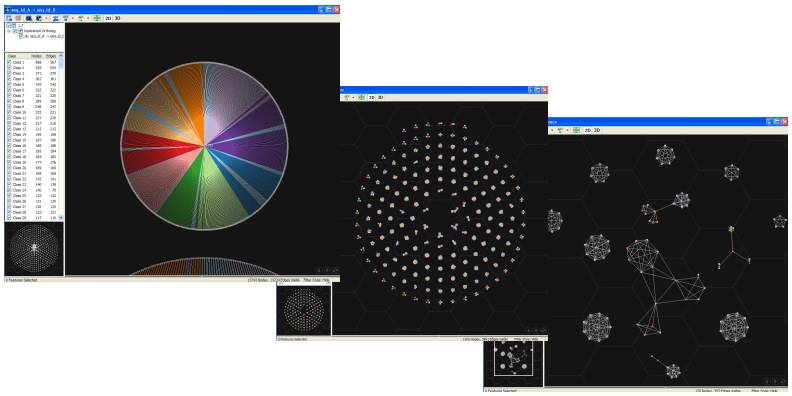
Challenges that May be Addressed with Advanced Computing and Mathematics Capabilities

- Defining the mechanism(s) of biotic and abiotic corrosion
- Understanding energy flow and how it relates to corrosion
- Designing strategies for ameliorating or reversing corrosion

Genomics and Metagenomics



Chris Oehmen
Sr. Research Scientist
Pacific Northwest
National Lab



High performance sequence analysis- throughput and complexity

- next-gen sequence data (reads →genes, genomes; experimental system)
- multiple genomes (*eg.* ortholog maps across *all* genomes)
- communities (predicting functions, pathways)
- visualization, hypothesis detection

Challenges that May be Addressed with Advanced Computing and Mathematics Capabilities

• community function models from pathway predictions→natural and engineered systems for energy and climate *but…*

- *How do you assemble a metagenome? (graph problem needing TB's memory)*
- *What is the minimal genome? (network of networks problem with millions of nodes)*
- *How do we find a hypothesis in all this data?*

- **Genomes**
- Sequence comparison
- Genome assembly/ annotation
- recombination models
- phylogenomic tools

- **Community structure/ dynamics**
- models of natural selection/ migration

- **Metabolites**
- interpret mass spectra
- metabolic reconstruction from genomes
- high-res (intracellular)/high-level (environmental) metabolic models

- **Physical/Biological environmental models**
- PDE- based
- diffusion/reaction
- flows
- fitness/replicators (Darwin project)

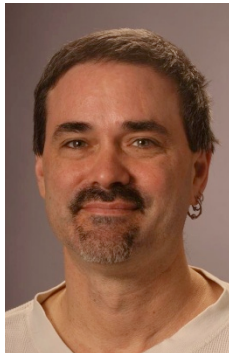
Inverse problem of "what metabolisms are active and where?"

quantitative ecosystem models

predictive dynamics

validate against transcriptomics/dynamics

Populations, Communities, Ecosystems and Evolutionary Dynamics: Genomics and Metagenomics



Owen White
Director of Bioinformatics
Institute for Genome Sciences,
University of Maryland
School of Medicine

Research Interests

- **NIH Human Metagenomic Project Coordination**
- **Genome Standards, interoperation**
- **Genomic/metagenomic annotation systems**

Challenges that May be Addressed with Advanced Computing and Mathematics Capabilities

- **VM-based genomic and metagenomic analysis systems**
- **Data-intensive cloud computing**
- **Consensus annotation**